STOCHASTIC APPROXIMATION: THEORY AND APPLICATIONS

Dr. Gugan Thope

Department of Computer Science and Engineering

**Indian Institute of Science, Bangalore**

Week 7

Lecture 28

Convergence Rate for Linear Stochastic Approximation - Part 2

Hello and Namaste, everyone. Welcome to Lecture 28 of this NPTEL course on Stochastic Approximation. Let us do a quick recap of what we have been doing so far. In the previous class, we started looking at the convergence rates of stochastic approximation algorithms. In particular, we wanted to ask,

What is the effect of step size on the convergence rates? For ease of discussion, we focused on linear stochastic approximation. In particular, we focused on this mean estimation algorithm. And then we showed that for the step size choice alpha_n equals 1 over n plus 1, the mean squared error—that is, the expected value of (x_n minus mu) the whole square—that quantity decayed at the rate of 1 over n. And keeping that in mind, we then asked this question: Okay, this convergence rate we got for the step size choice of 1 over n. What would have happened if we chose a more slowly decaying step size?

For example, what would have happened if we chose a step size choice of the form alpha_n equals 1 over (n plus 1) to the power alpha? Would we get a better convergence rate, or would we get a worse convergence rate? So, to derive the convergence rate under this setup, we derived an intermediate expression in the previous lecture, and in today's class, we will actually obtain a bound on this intermediate expression, right? And as we saw in the previous class—I mean, so that there is no suspense—the convergence rate with a slowly decaying step size was actually poorer, and we will derive this today. At the end, we will discuss some consequences of this convergence rate. With this in place, let us do a formal recap and let us continue our discussion from there.

So, last time we looked at this specific update rule of the form: little xn plus 1 is little xn plus alpha n times capital Xn plus 1 minus Xn. So, as I said, this is referred to as the mean estimation algorithm. Since this algorithm can be used to estimate the expected value of the random variable that is stated over here, and in the previous class, we showed that if alpha n is 1 over n plus 1, then the mean squared error is sigma square over n. Our goal then was to ask what would happen to the convergence rate if we chose a slowly decaying step size of this form—that is, a step size of the form 1 over n plus 1 to the power alpha, where alpha is a number between 0 and 1, right? And we define theta n to be X n minus mu and then we derive this intermediate expression: the expected value of theta n square is sigma square sum of k equals 0 to n minus 1 alpha k square times the product of j equals k plus 1 to n minus 1 of 1 minus alpha j square.
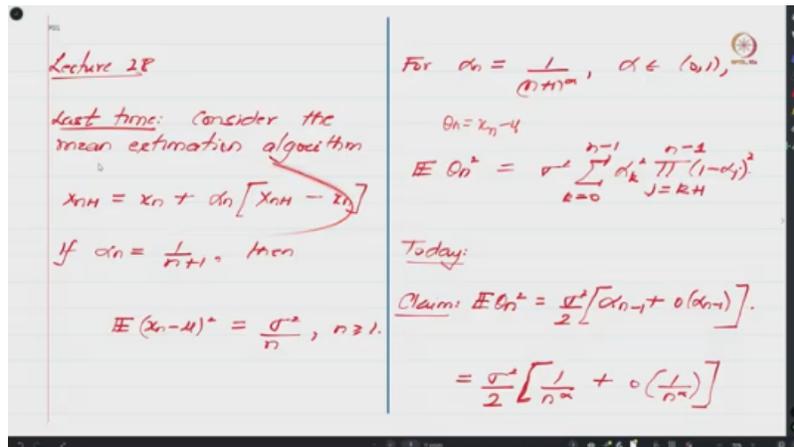
And in today's class, what we will do is derive an upper bound on this expression so that we can show that the expected value of theta n square is sigma square over 2 times alpha n minus 1 plus an expression that is little o of alpha n minus 1, which means that for sufficiently large n, this expression dominates the sum. Hence, this expression dominates the convergence rate of Xn. And if you substitute alpha n minus 1 as defined over here, then this expression will lead to 1 over n to the power alpha, and this expression will be little o of 1 over n to the power alpha, again confirming that this is the expression that will dominate the sum. Hence, the final convergence rate in a mean squared error sense will be of the order of 1 over n to the power alpha, which is worse than the order 1 over n that we had seen over here. Right.
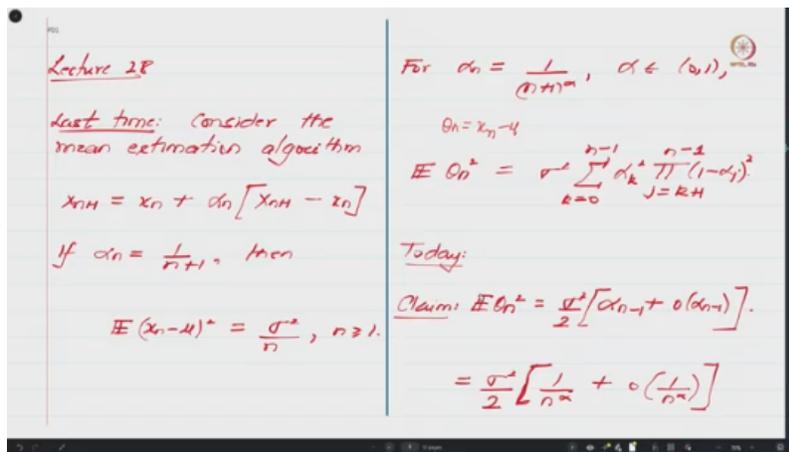
$$x_{n+1} = x_n + \alpha_n\left[X_{n+1} - x_n\right]$$

$$\alpha_n = \frac{1}{n+1}$$

$$E\left(x_n - \mu\right)^2 = \frac{\sigma^2}{n}, \; n \geq 1$$

$$\alpha_n = \frac{1}{(n+1)^\alpha}, \; \alpha \in (0, 1),$$

$$E\theta_n^2 = \sigma^2 \sum_{k=0}^{n-1} \alpha_k^2 \prod_{j=k+1}^{n-1} \left(1 - \alpha_j\right)^2$$

So, let us first formally verify this claim.

So, what we are going to do is take whatever is over here—right—we are going to define it as Sn, as I have stated over here, right? And then, let us denote this product that we have over here as Uk n, that is, the product of 1 minus alpha j to the power—I mean, the whole square—from j equals k plus 1 to n minus 1. Again, whenever we have a product where the lower index is larger than the upper index, we will interpret this product as being 1. So, with this notation, one can see that Sn has the form summation k equals 0 to n minus 1 of alpha k square. So, whatever we have over here, I have written here, and this expression we have denoted as ukn, and hence I have stated it as ukn. Right.

Proof: Let
$$S_n = \sum_{k=0}^{n-1} \alpha_k^2 \prod_{j=k+1}^{n-1} (1-\alpha_j)^2$$

Further, let
$$u_k^{(n)} = \prod_{j=k+1}^{n-1} (1-\alpha_j)^2$$

Then, $S_n = \sum_{k=0}^{n-1} \alpha_k^2 u_k^{(n)}$

We first obtain a bound on $u_k^{(n)}$.

We have
$$u_{k-1}^{(n)} = (1-\alpha_k)^2 u_k^{(n)}$$

Hence,
$$u_k^{(n)} - u_{k-1}^{(n)} = u_k - (1-\alpha_k)^2 u_k^{(n)}$$
$$= [2\alpha_k - \alpha_k^2] u_k^{(n)}$$

So, what we will do is rewrite this expression so that the $\alpha_{\square-1}$ term comes out cleanly and whatever remains. Right. We will show that the remainder is little-o of $\alpha_{\square-1}$. So, toward that, we are going to first rewrite this expression, and toward that, what we will do is make an observation. The first observation is that if you take this expression starting from k, which is when you would get this $u_{\square-1}$, all the way up to n.

So, then this expression is equal to $(1 - \alpha_{\square}^2) \times u_{\square\square}$, right? $u_{\square\square}$ starts from k + 1 to n - 1, whereas $u_{\square-1}$ starts from k to n - 1. Hence, $u_{\square-1}$ has this relation with $u_{\square\square}$, right? So, from this, I hope you agree that if I take the difference between $u_{\square\square}$ and $u_{\square-1\square}$, we would end up with—so I should add superscript n here. So, we would end up with this expression as is. And this expression is $(1 - \alpha_{\square})^2 \times u_{\square\square}$, and hence, if I pull $u_{\square\square}$ out as common, I will be left with $1 - (1 - \alpha_{\square})^2$, which leads to the expression here: $2\alpha_{\square} - \alpha_{\square}^2$.

Is this okay? And from this, if I pull $\alpha_{\square}$ out as common, one can see that we would end up with $\alpha_{\square} \times (2 - \alpha_{\square}) u_{\square\square}$, right? So, what I will do is take this expression to the left-hand side, right? And multiply both sides by $\alpha_{\square}$ so that we can see that $\alpha_{\square}^2 u_{\square\square}$—that is, this term with α— $\alpha_{\square}$ multiplied on the right-hand side is $\alpha_{\square} / (2 - \alpha_{\square})$, which comes by moving it to the other side.

Left panel:

$$= \alpha_k (2 - \alpha_k) u_k^{(n)}$$

Therefore,

$$\alpha_k^2 u_k^{(n)} = \left(\frac{\alpha_k}{2 - \alpha_k}\right)\left[u_k^{(n)} - u_{k-1}^{(n)}\right];$$

note that

$$\alpha_k < 2$$

and, hence, $2 - \alpha_k > 0, \ \forall k \geq 0$

Right panel:

Let $g_k = \dfrac{\alpha_k}{2 - \alpha_k}$ for $k \geq 0$

and 0 for $k = -1$.

Then,

$$S_n = \sum_{k=0}^{n-1} \alpha_k^2 \, u_k^{(n)}$$

$$= \sum_{k=0}^{n-1} g_k \left[u_k^{(n)} - u_{k-1}^{(n)}\right].$$

So, this expression times whatever was there on the left-hand side. So, the expression on the left-hand side was u_k^n minus u_{k-1}^n, right? And hence, one can see that this expression equals this, right? The advantage of working with this instead of this is that we have this difference, and I will soon show you how this difference actually helps. Now, you know, whenever we write some ratio in this form, one of the key things we have to ensure is that the denominator is strictly positive. That is indeed true in our case because α_k is less than or equal to 1 for any k, and hence α_k is strictly less than 2 for all values of k, which implies that the denominator is strictly positive for all values of k. So, now if we define G_k to be this expression, then the expression for S_N that we had started out with can be rewritten as the sum from k equals 0 to N minus 1 of G_k times the difference of the successive U_k terms.

All right. So, as I told you, because of the presence of successive difference terms, there is some advantage, and I will soon tell you what that advantage is, right? So, S_N equals this, this equals this. This is just a restatement of what we observed previously. So, now let us substitute different values of k equals 0 over here, right?

Since

$$S_n = \sum_{k=0}^{n-1} \alpha_k^2 \, u_k^{(n)}$$

$$= \sum_{k=0}^{n-1} g_k \left(u_k^{(n)} - u_{k-1}^{(n)} \right)$$

$$= g_0 \left(u_0^{(n)} - u_{-1}^{(n)} \right) + g_1 \left(u_1^{(n)} - u_0^{(n)} \right)$$

$$+ g_2 \left(u_2^{(n)} - u_1^{(n)} \right) + \cdots +$$

$$g_{n-1} \left(u_{n-1}^{(n)} - u_{n-2}^{(n)} \right).$$

$$= g_{n-1} \, u_{n-1}^{(n)} - g_0 \, u_{-1}^{(n)}$$

$$+ (g_0 - g_1) u_0^{(n)}$$

$$+ \cdots$$

$$+ (g_{n-2} - g_{n-1}) u_{n-2}^{(n)}$$

$$= g_{n-1} \, u_{n-1}^{(n)} - g_0 \, u_{-1}^{(n)}$$

$$+ \sum_{k=0}^{n-2} (g_k - g_{k+1}) u_k^{(n)}$$

Now, $u_{n-1}^{(n)} = \prod_{j=n}^{n-1} (1 - \alpha_j)^2 = 1.$

So, by writing different values of k equals 0, the first term that we will end up with is G_0 times u_0^n minus u_{-1}^n. The second term that we will have would be G_1 times u_1^n minus u_0^n, all the way until the last term, which will be G_{n-1} (obtained by substituting k equals n minus 1) times u_{n-1}^n minus u_{n-2}^n. So, now what we will do is we will take this difference over here and rewrite it in terms of differences of G_k, and towards that, what we will do is we will collect common terms. So, you can see that, you know, here you have this u_0^n, and here also you have u_0^n. So, we will combine them so that we end up with G_0 minus G_1, which is what we have over here. And in this way, we go all the way up till u_{n-2}^n, which will end up giving us the expression G_{n-2} minus G_{n-1} times u_{n-2}^n. So, after we do this, we will be left with this expression and this product.

So, which I have written separately over here. So, this is g n minus 1 times u n minus 1 all the way up till n minus g 0 times u minus 1 n. This is the expression that we have over here. So, what we will do is we will, you know, write this sum of terms in shorthand form so that it helps with our analysis, right? So, all the terms that are present over here can be written as sum equals k equals 0 to n minus 2. Of, you know, product of two terms where the first term is the difference between successive terms g 0 minus g 1, g n minus 2 minus g n minus 1. In that way, one can see that this general term will be g k minus g k plus 1, and the expression over here is u k n and k goes from 0 to n minus 2.

Is this okay? So now what we will do is we will get some simplified expression for this and show that this is the expression that leads to the dominating term in the convergence

rate analysis, whereas this expression is little o of alpha n minus 1. So this is what the rest of the analysis is, right? And to begin with, So, let us, you know, look at this expression over here. So, if you look at u n minus 1 of n, right.

So, this will start from j equals n and go all the way up till n minus 1, this product of 1 minus alpha j square. So, as I told you, whenever the index is lower index is larger than the upper index, we will interpret this product to be 1. Hence, this term which appears over here will be 1. So, this expression will basically be g n minus 1, and on the other hand, if you look at g minus 1, right, g minus 1 we had defined it to be 0. So, I mean, I think I forgot to mention it. So, remember I said that we define g k to be alpha k over 2 minus alpha k, right? And you know, if you define g of x to be like 1 over x plus 1 to the power alpha by 2 minus 1 over x plus 1 to the power alpha.

And if you take, you know, your X to actually minus 1, one can see that this expression actually goes to 0, right? And keeping that in mind, we define this value of GK for K equals minus 1 as 0, right. So, if you keep that interpretation in mind, right, we end up observing that g minus 1 is 0, which is the term that is—sorry, I think I have made a mistake, so let me just confirm this. Sorry, I should say, so this is the term g0 u minus 1, and what is g0? g0 is this expression over here, and this is alpha 0 by u minus alpha 0.

Page 1 (top image):

Left column:
$$= \alpha_k(2 - \alpha_k)\, u_k^{(n)}$$

Therefore,

$$\alpha_k^{\perp}\, u_k^{(n)} = \left(\frac{\alpha_k}{2 - \alpha_k}\right)\left[u_k^{(n)} - u_{k-1}^{(n)}\right];$$

note that

$$\alpha_k < 2$$

and, hence, $2 - \alpha_k > 0$, $\forall k \geq 0$

Right column:
Let $g_k = \dfrac{\alpha_k}{2 - \alpha_k}$ for $k \geq 0$

and 0 for $k = -1$.

Then,

$$g(z) = \frac{\frac{1}{(z+1)^\alpha}}{2 - \frac{1}{(z+1)^\alpha}}$$

$$S_n = \sum_{k=0}^{n-1} \alpha_k^{\perp}\, u_k^{(n)}$$

$$= \sum_{k=0}^{n-1} g_k\left[u_k^{(n)} - u_{k-1}^{(n)}\right].$$

Page 2 (middle image) — duplicate of page 1:

Page 3 (bottom image):

Left column:
Further,

$$g_{-1} = 0$$

Hence,

$$S_n = g_{n-1} + \sum_{k=0}^{n-1}(g_k - g_{k+1})\, u_k^{(n)}$$

Now, $g_{n-1} = \dfrac{\alpha_{n-1}}{2 - \alpha_{n-1}}$

$$= \frac{\alpha_{n-1}}{2} + \frac{\alpha_{n-1}}{2 - \alpha_{n-1}}.$$

Right column:
We now show that

$$\sum_{k=0}^{n-1}(g_k - g_{k+1})\, u_k^{(n)} = o(\alpha_{n-1})$$

$$g_k = \frac{\alpha_k}{2 - \alpha_k} \quad \& \quad g_{k+1} = \frac{\alpha_{k+1}}{2 - \alpha_{k+1}}$$

Hence, if we define

$$g(z) = \frac{z}{2 - z},$$

So, I think I have made a mistake here. Just a minute, let me quickly fix this. Okay, so I think I have made a mistake here. This should be—I mean, while g minus 1 is 0, this expression is what we need to worry about. But since there is a negative sign over here,

okay, we can actually ignore this term and, you know, say that this expression is upper bounded, right? by this quantity over here plus the sum of these terms, okay.

So, this expression—all I have shown is that this expression is 1, right? And hence, we would end up with g n minus 1 plus the sum of the remainder terms, which I write as is over here. Is this okay? So, here actually I should have—so here I have n minus 2. So, let me just make sure this is n minus 2.

So, I start from 0 to n minus 2. So, I think I should have n minus 2 over here, and now with this in place, let us do the analysis. So, now let us first look at this quantity over here. Now, this quantity, as I told you, is alpha n minus 1 divided by 2 minus alpha n minus 1. This is how we had defined your GK expression over here.

And hence, now what we will do is rewrite this expression by adding and subtracting alpha n minus 1 over 2, okay. So, I can add and subtract alpha n minus 1 over 2, right? So, if I add and subtract this quantity, it will lead to the first term over here, and here, wherever I have alpha n minus 1 over 2 minus alpha n minus 1. So, this expression, I will subtract it from this expression.

So, if I take alpha n minus 1 out in common, what I will end up with over here is alpha n minus 1, 1 over 2 minus alpha n minus 1 minus half. So, if I multiply the denominators and put 2 over here and 2 minus alpha n minus 1 over here and take their subtraction, I would end up with alpha n minus 1 in the numerator as well. This is why I have alpha n minus 1 squared over here, and this term and the 2 over here will actually lead me to 2 times 2 minus alpha n minus 1. Sorry, there are a couple of typos here. I am sorry about that.

So, now observe that this expression that we have is the dominant term that we wanted in the convergence rate analysis. And as your n becomes larger and larger, this term that we have goes to 0, right, and hence the numerator is what will dictate the order at which this expression actually decays. In other words, this denominator, okay, can be shown to be lower bounded by a constant for large enough values of n, and hence this whole expression is upper bounded by some constant times alpha n minus 1 squared, right, and hence this expression can be shown to be little o of alpha n minus 1. So, as a first step, we have shown that Sn is upper bounded by this sum, and the first term over here we have

shown that, you know, it is having alpha n minus 1 over 2 as the dominant term, and the remaining term is little o of alpha n minus 1.

In the remainder of this analysis, what we will show is this expression is actually little o of alpha n minus 1 over here. So, this is what we are going to show. So, actually, I have, you know, worked with n minus 1 over here because of the mistake, but one can, you know, do this analysis by whatever I am going to discuss by presuming n-2 over here; it does not change anything, right, and one can also see that this expression over here is upper bounded by this because I am adding a term over here, right, and it does not matter, okay. So, keep that thing in mind, and let us do the analysis over here, right.

So, the first goal that we have right now is to show what can we say about this gk minus gk plus 1. Now, as you remember, gk is of the form alpha k over 2 minus alpha k, and gk plus 1 is alpha k plus 1 minus 2 over 2 minus alpha k plus 1. Hence, if we define g of x to be of the form x over 2 minus x. Then this quantity is g of alpha k, and this quantity that we have over here is g of alpha k plus 1. In other words, if you define this function from R to R and you evaluate this function over here at x equals alpha k, then we would end up with gk, and if you evaluate this function at alpha k plus 1, then whatever you end up with will be your gk plus 1.

And hence I have written gk is g of alpha k and hence if you invoke the mean value theorem one can now conclude that gk minus gk plus 1 over alpha k minus alpha k plus 1 which is exactly g of alpha k minus g of alpha k plus 1 divided by alpha k minus alpha k plus 1. So, this expression must equal the derivative of g at some intermediate value between alpha k and alpha k plus 1. So, this is you know something very well known because your g of x actually is a well defined function between whenever x lies between 0 and 1. right and from this fact what one can see is that you know this g prime c that we have over here to in order to bound this let us first compute what is g prime x right a simple calculus shows that since g of x is, sorry, since g of x is, let me just make sure I am doing it correctly.
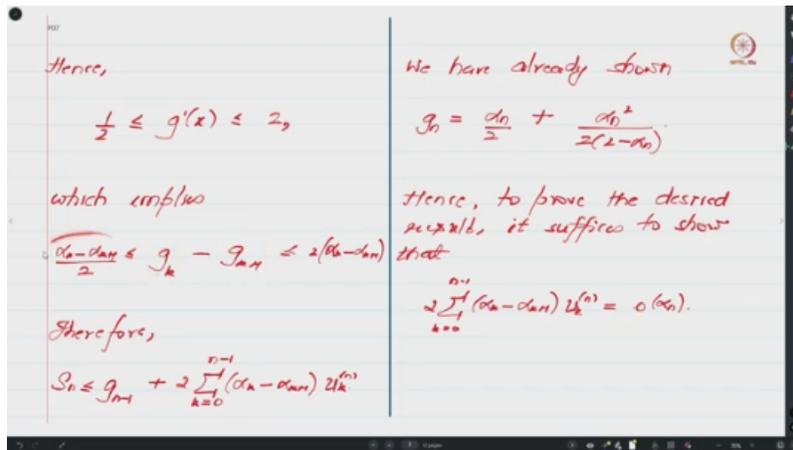
then

$$g_k = g(\alpha_k) \text{ for } k \geq 0$$

Hence, from the mean value theorem,

$$\frac{g_k - g_{k+1}}{\alpha_k - \alpha_{k+1}}$$

$$= \frac{g(\alpha_k) - g(\alpha_{k+1})}{\alpha_k - \alpha_{k+1}} = g'(c)$$

for some $c \in (\alpha_k, \alpha_{k+1})$.

Since $g'(x)$

$$= \frac{1}{2-x} + \frac{x}{(2-x)^2}$$

$$= \frac{2}{(2-x)^2},$$

we have $g'(x) \in \left[\frac{1}{2}, 2\right]$,

since $0 \leq x \leq 1$.

Since g of x is actually x by 2 minus x, the derivative of this expression will indeed be this one can, you know easily verify this. And you know if I multiply and divide the first expression by 2 minus x on the numerator and denominator, one can see that the 2 minus x plus x will lead to a 2 and the denominator results in 2 minus x the whole square. And we are going to focus on the case where alpha k actually takes values between 0 and 1 and wherever there was x actually we have replaced it with alpha x. And hence, our x also lies between 0 and 1, right? And this tells us that your g prime x actually takes values between half and 2, right?


then

$$g_k = g(\alpha_k) \text{ for } k \geq 0$$

Hence, from the mean value theorem,

$$\frac{g_k - g_{k+1}}{\alpha_k - \alpha_{k+1}}$$

$$= \frac{g(\alpha_k) - g(\alpha_{k+1})}{\alpha_k - \alpha_{k+1}} = g'(c)$$

for some $c \in (\alpha_k, \alpha_{k+1})$.

Since $g'(x)$

$$g(x) = \frac{x}{2-x}$$

$$= \frac{1}{2-x} + \frac{x}{(2-x)^2}$$

$$= \frac{2}{(2-x)^2}, \qquad 0 \leq \alpha_k \leq 1$$

we have $g'(x) \in \left[\frac{1}{2}, 2\right]$,

since $0 \leq x \leq 1$.

So, you can use these bounds on x to conclude that g prime x will also take values between half and 2. So, from this expression, if we put an upper bound of you know 2 over here, one can conclude that your g prime x actually lies between half and 2 and hence one can conclude that gk minus gk plus 1 is on the one hand lower bounded by

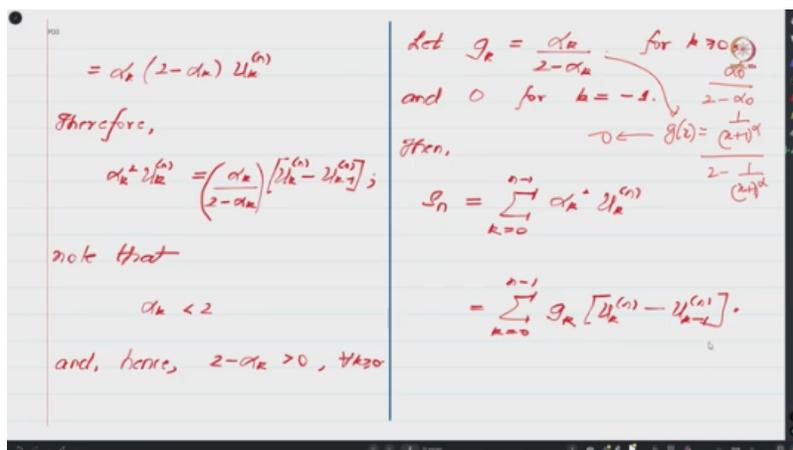alpha k minus alpha k plus 1 over 2 and on the other hand, it is upper bounded by 2 times alpha k minus alpha k plus 1. Therefore, one can conclude that Sn is upper bounded by Gn minus 1 which is the term that we had over here, right? It is upper bounded by this plus 2 times wherever you had Gk minus Gk plus 1, we can now replace it by alpha k minus alpha k plus 1 and put a 2 outside so that this expression now is an upper bound for Sn, right? right and you know now what we will okay so I should say this is like alpha n-1 here so we have already shown that g n-1 is alpha n-1 alpha n-1 alpha n-1 so this term

Hence,

$$\frac{1}{2} \le g'(x) \le 2,$$

which implies

$$\frac{\alpha_k - \alpha_{k+1}}{2} \le g_k - g_{k+1} \le 2(\alpha_k - \alpha_{k+1})$$

Therefore,

$$S_n \le g_{n-1} + 2\sum_{k=0}^{n-1}(\alpha_k - \alpha_{k+1}) u_k^{(n)}$$

We have already shown

$$g_n = \frac{\alpha_n}{2} + \frac{\alpha_n^2}{2(1-\alpha_n)}.$$

Hence, to prove the desired result, it suffices to show that

$$2\sum_{k=1}^{n-1}(\alpha_k - \alpha_{k+1}) u_k^{(n)} = o(\alpha_n).$$
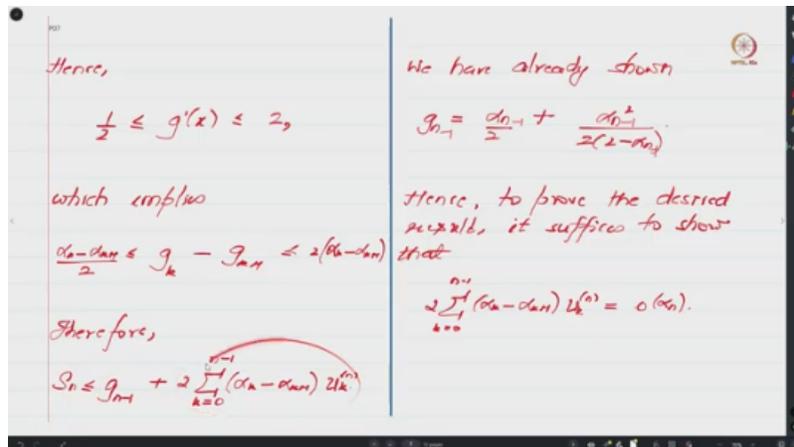
We have already shown it is like of the order alpha n minus 1, with this term decaying quite fast. Hence, to prove the desired result, it suffices to show that whatever the remainder term is, which is what we have written over here, actually decays very, very fast. In particular, it is little o of alpha n. So, this is what remains to be shown, right? So, let me quickly summarize because this proof involves a bit of algebra. What we have done is, right?

Hence,

$$\frac{1}{2} \le g'(x) \le 2,$$

which implies

$$\frac{\alpha_k - \alpha_{k+1}}{2} \le g_k - g_{k+1} \le 2(\alpha_k - \alpha_{k+1})$$

Therefore,

$$S_n \le g_{n-1} + 2\sum_{k=0}^{n-1}(\alpha_k - \alpha_{k+1}) u_k^{(n)}$$

We have already shown

$$g_{n-1} = \frac{\alpha_{n-1}}{2} + \frac{\alpha_{n-1}^2}{2(2-\alpha_{n-1})}.$$

Hence, to prove the desired result, it suffices to show that

$$2\sum_{k=0}^{n-1}(\alpha_k - \alpha_{k+1}) u_k^{(n)} = o(\alpha_n).$$
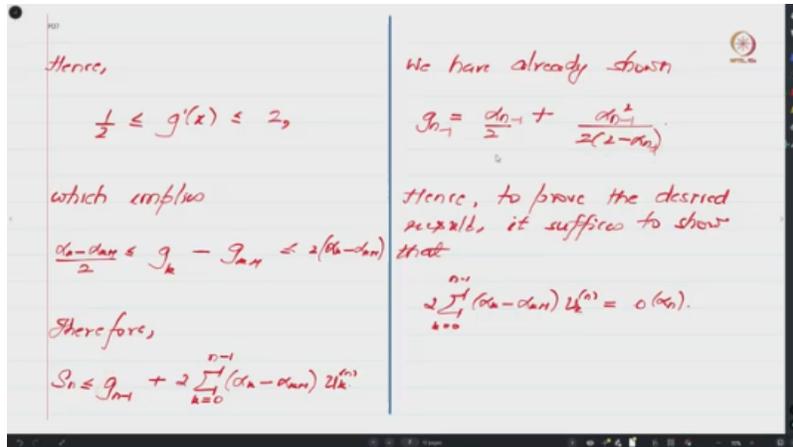
We started off with an expression of this form, and we took this expression and wrote it in this form. Thereafter, what we did was rewrite this expression so that instead of having differences between UKs, we have differences between GKs. So, we have an expression of this form. And now what we are doing is using the intermediate value theorem to replace this gk minus gk plus 1 with alpha k minus alpha k plus 1, which is what we have done. So, if you notice, we have said gk minus gk plus 1 is upper bounded by 2 times alpha k minus alpha k plus 1, and hence we have obtained this expression. So, in some sense, we have done a bit of algebra so that we can express Sn in terms of Gn-1 plus some expression. This term we have already shown to have the dominant term alpha n-1 over 2, and hence now it remains to show that this secondary term is also of lower order to conclude the final result.

$$= \alpha_k(2-\alpha_k) u_k^{(n)}$$

Therefore,

$$\alpha_k^2 u_k^{(n)} = \left(\frac{\alpha_k}{2-\alpha_k}\right)\left[u_k^{(n)} - u_{k-1}^{(n)}\right];$$

note that

$$\alpha_k < 2$$

and, hence, $2-\alpha_k > 0$, $\forall k \ge 0$

Let $g_k = \frac{\alpha_k}{2-\alpha_k}$ for $k \ge 0$

and 0 for $k = -1$.

Then,

$$S_n = \sum_{k=0}^{n-1} \alpha_k^2 u_k^{(n)}$$

$$= \sum_{k=0}^{n-1} g_k \left[u_k^{(n)} - u_{k-1}^{(n)}\right].$$

$\frac{\alpha_0}{2-\alpha_0}$

$0 \leftarrow g(k) = \frac{1}{(k+1)^q}$

$\frac{2-\frac{1}{(k+1)^q}}{(k+1)^q}$

Hence,

$$\frac{1}{2} \le g'(x) \le 2,$$

which implies

$$\frac{\alpha_n - \alpha_{n+1}}{2} \le g_k - g_{n+1} \le 2(\alpha_n - \alpha_{n+1})$$

Therefore,

$$S_n \le g_{n+1} + 2 \sum_{k=0}^{n-1} (\alpha_n - \alpha_{n+1}) u_k^{(n)}$$

We have already shown

$$g_{n+1} = \frac{\alpha_{n+1}}{2} + \frac{\alpha_{n+1}^2}{2(1-\alpha_{n+1})}.$$

Hence, to prove the desired result, it suffices to show that

$$2 \sum_{k=0}^{n-1} (\alpha_n - \alpha_{n+1}) u_k^{(n)} = o(\alpha_n).$$

So, let us see how to derive a bound on this expression. Again, this bound will involve a bit of algebra, but you will soon see that These are the kinds of analyses one needs to do or often ends up doing in order to derive convergence rates of stochastic approximation algorithms. Hence, I am taking you through this discussion so that you also get exposed to the kind of algebra that is typically involved in deriving such convergence rates. So, before I embark on the details, let me sort of give you a broad overview. Whatever is over here, I will sort of split it up into two parts.

The first n over 2 terms we will collect over here, and the second n over 2 terms in this sum involving n terms, we will collect in the second sum. In the first n over 2 terms, what we will do is we will, you know, obtain a bound on this ukn and use that bound to show that the first n over 2 terms decay quite fast, and the second n over 2 terms we will separately show that that expression also decays very fast. So, let us begin with this analysis. So, first, let us obtain an upper bound on ukn. Now, ukn recall is the product of terms of the form 1 minus alpha j square, where j goes from k plus 1 to n minus 1, and 1 minus alpha j square is upper bounded by e raised to minus 2 alpha j. Why is that?

Hence,

$$\frac{1}{2} \leq g'(x) \leq 2,$$

which implies

$$\frac{\alpha_n - \alpha_{n+1}}{2} \leq g_k - g_{n+1} \leq 2(\alpha_n - \alpha_{n+1})$$

Therefore,

$$S_n \leq g_{n+1} + 2 \sum_{k=0}^{n-1} (\alpha_n - \alpha_{n+1}) u_k^{(n)}$$

We have already shown

$$g_{n+1} = \frac{\alpha_{n+1}}{2} + \frac{\alpha_{n+1}^2}{2(2-\alpha_n)}$$

Hence, to prove the desired results, it suffices to show that

$$2\sum_{k=0}^{n-1} (\alpha_n - \alpha_{n+1})(u_k^{(n)}) = o(\alpha_n).$$

$$\frac{n}{2} \qquad\qquad \frac{n}{2}$$

Now,

$$u_k^{(n)} = \prod_{j=k+1}^{n-1} (1-\alpha_j)^2$$

$$\leq e^{-2\sum_{j=k+1}^{n-1} \alpha_j},$$

which follows since

$$1-x \leq e^{-x}$$

Hence, $u_k^{(n)} \leq e^{-2(n-1-k)\alpha_{n-1}}$

For $k \leq \lfloor \frac{n}{2} \rfloor - 1 \leq \frac{n}{2} - 1$,

$$u_k^{(n)} \leq e^{-2\frac{n}{2}\cdot \alpha_{n-1}}$$

$$= e^{-n\frac{1}{n^\alpha}} = e^{-n^{1-\alpha}}$$

Well, recall that 1 minus x is less than e raised to minus x. And because this is true, 1 minus x square is less than e raised to minus 2x whenever x is greater than or equal to 0, which is indeed the case for us because alpha j is bigger than or equal to 0. And hence, one can conclude that this 1 minus alpha j square, this particular term, is upper bounded by e raised to minus 2 alpha j. This product over here is upper bounded by e raised to minus 2 sum of j equals k plus 1 to n minus 1 of alpha j. So, this is something that is immediate to see. And then what we can see is that since your alpha j's are decaying, the smallest term over here will be the last term, which is alpha n minus 1.

Now,

$$u_k^{(n)} = \prod_{j=k+1}^{n-1}(1-\alpha_j)^2$$

$$\leq e^{-2\sum_{j=k+1}^{n-1}\alpha_j},$$

which follows since

$$1-x \leq e^{-x}$$

Hence, $u_k^{(n)} \leq e^{-2(n-1-k)\alpha_{n-1}}$

For $k \leq \lfloor\frac{n}{2}\rfloor - 1 \leq \frac{n}{2}-1$,

$$u_k^{(n)} \leq e^{-2\frac{n}{2}\cdot\alpha_{n-1}}$$

$$= e^{-n\frac{1}{n^\alpha}} = e^{-n^{1-\alpha}}$$
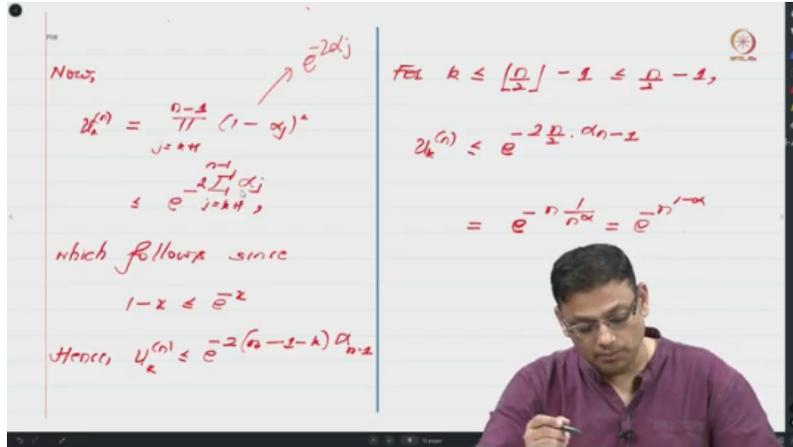
And because of the negative sign, one can show that this quantity over here is lower bounded by this expression. In particular, this sum that we have here. Right, sum of j equals k plus 1 till n minus 1 alpha j can be shown to be bigger than n minus 1 minus k, which basically corresponds to how many terms we have in this sum, times alpha n minus 1, which is the smallest of the expressions over here. So, this is exactly what we have, and hence I have replaced this sum with a lower bound, which ends up giving us this upper bound e raised to minus 2 of n minus 1 minus k alpha n minus 1.

$$\sum_{j=k+1}^{n-1}\alpha_j \geq (n-1-k)\alpha_{n-1}$$
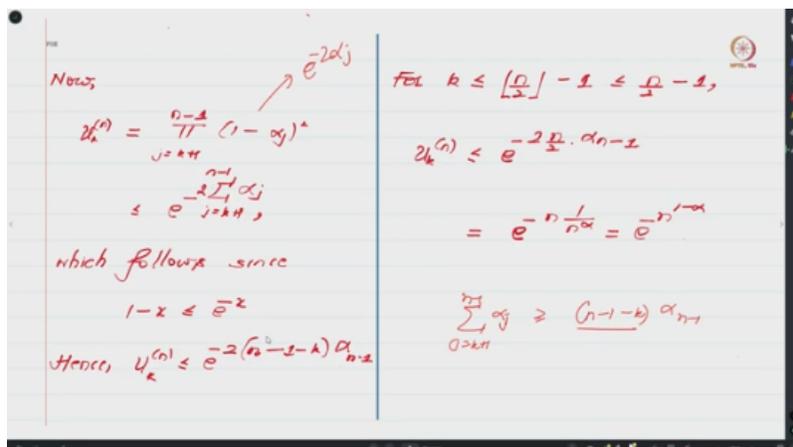
So, a lower bound on this quantity translates into an upper bound on this quantity because of the presence of this negative sign. So, now what we will do is we will work on the case where k is roughly less than n over 2 minus 1 and then see if we have a clean expression for this upper bound on ukn. So, ukn is some expression like this, and here you can see that. This is like e raised to 2 n minus 1, right, minus k, and since your k is less than n

over 2 minus 1, this expression—one can show—is n minus 1. n minus 1 minus k is greater than, since k is less than this, this is greater than n minus 1. Minus n over 2 minus 1, which is like n over 2.

So, just check this calculation. I may have made a minor typo possibly, but the spirit of the calculation should go through. So, whenever k is less than or equal to this, one can show that. You know, this n minus 1 minus k is actually lower bounded by n minus 1 minus n over 2 minus 1 with round brackets over here, which will lead to n over 2. Hence, this ukn expression is upper bounded by e raised to minus 2 times whatever you had over here. This is.

Lower bounded by n over 2; hence, you have n over 2 and whatever alpha n minus 1. I write it as it is, so these twos cancel off, and I end up with e raised to minus n times alpha n minus 1. And your alpha n recall has this expression, you know, 1 over. n plus 1 raised to alpha. So, if I substitute n minus 1 in place of n, I would end up with 1 over n to the power alpha, and hence if I pull it up in the numerator, I will end up with e raised to minus n raised to 1 minus alpha over here. Is this okay? Now, observe that whenever alpha is strictly less than 1, this term will be positive, which implies that this ukn expression actually decays at an exponential rate, right?

So, this ukn expression actually decays at an exponential rate. Hence, if you look at roughly the first n over terms in that second term in the upper bound on Sn, one can see that this expression can be upper bounded by, notice that for all k, right? Let me just emphasize this: this bound over here is true for all k, which is less than the, you know,

floor of n over 2 minus 1, right? Like roughly n over 2 terms, right? So, for all k which is less than roughly n over 2 index This ukn is upper bounded by this, and notice that there is no k over here, which implies I can uniformly bound this ukn with this expression and retain the other remaining terms. We have something like this.

And one can again, you know, write floor of this over here, right? And this then leads up to a telescopic sum. And because we have a telescopic sum, I can add and subtract similar terms from subsequent terms. And hence, we would end up with, you know, a term like alpha 0, which comes from k equals 0 and some last expression, which is like, you know, n over 2. So, it should end up with some expression like this.

So, we will end up with something like this, and this quantity over here is positive. Hence, I can drop this whole thing from this expression, and we know that alpha 0 is 1, and hence this expression is upper bounded. by e raised to minus n minus 1 minus alpha,

which we get from here. And since this is exponentially decaying, one can trivially check that this expression is big O of 1 over n, that is, this expression decays at the rate of 1 over n. So, now what remains is to, you know, look at terms of this form but for k bigger than or equal to floor of, so I should say k greater than or equal to n over floor of n over 2.

That means you know the previous terms we have considered in this sum, and the terms that remain we are now going to discuss. So, we have some expression like this where K is strictly bigger than or equal to—I mean, I should not say strictly—it is bigger than or equal to the floor of n over 2. And now what we will do is, in the previous case, we sort of focused on getting a bound on uK n. In this case, that is for K bigger than or equal to the floor of n over 2, what we will do is we will try to obtain a bound on the difference between alpha K and alpha K plus 1, which appears over here. Right, and let's say alpha k is, you know, h evaluated at k. Recall that h of x, if we define it as 1 over x plus 1 to the power alpha, then alpha k is h of k because alpha k, recall, is 1 over k plus 1 to the power alpha.

So, this is the expression that we have. Hence, if I define h of x in this fashion, in place of x, if I put k, then I would end up with alpha k over here. And similarly, alpha k plus 1 is h of k plus 1. Again, these are all well-defined quantities which are continuous and differentiable when k is finite. And hence, one can make use of

Again, the mean value theorem to conclude that h of k minus h of k plus 1 will equal h prime c times the difference between k and k plus 1, which is minus 1. Now, if I look at

this quantity and take its derivative, right? So, the derivative of this expression is minus alpha over x plus 1 raised to alpha minus 1, right? Or otherwise, let me just make sure I have done it correctly. So, this is minus alpha x plus 1 minus alpha minus 1, which is like minus alpha x plus 1.

x plus 1 raise to 1 plus alpha which is exactly what I have written over here right. So, H prime C is minus alpha that minus alpha and this minus 1 make it a plus hence we end up with alpha times C plus 1 raise to minus 1 minus alpha. right and you know now this C actually lies between K and K plus 1 and since this quantity over here is negative right the largest value over here will be obtained for C equals K and hence one can see that this whole expression is actually upper bounded okay so I think this C is not I mean I didn't mean to use a C over here let me maybe call this as some constant So, let us say capital K, K also I should not use. So, what other constant?

So, let us say I call it as A. So, capital A which is some constant times this expression over here. So, this constant K, I think I can actually use alpha itself. So, let me use alpha here. So, this is alpha times K plus 1 raised to minus 1 minus alpha. what I do is I ask for K which is bigger than equal to floor of n over 2 okay what is an upper bound so K is larger so since K is sufficiently large right I can say that this expression is upper bounded by some constant times the smallest value that K can take now notice that the smallest value that K can take is

Hence,

$$\sum_{k=0}^{\lfloor n/2 \rfloor - 1} (\alpha_k - \alpha_{k+1})\, u_k^{(n)}$$

$$\le e^{-n^{1-\alpha}} \sum_{k=0}^{\lfloor n/2 \rfloor - 1} (\alpha_k - \alpha_{k+1})$$

$$= e^{-n^{1-\alpha}} \left[\alpha_0 - \alpha_{\lfloor n/2 \rfloor + 1}\right]$$

$$\le e^{-n^{1-\alpha}} = O\!\left(\tfrac{1}{n}\right)$$

On the other hand, for $k \ge \lfloor \tfrac{n}{2} \rfloor$,

if we let $h(x) = \dfrac{1}{(x+1)^\alpha}$,

$$\alpha_k = \dfrac{1}{(k+1)^\alpha}$$

then $\alpha_k - \alpha_{k+1}$

$$= h(k) - h(k+1) \qquad -\alpha (x+1)^{-\alpha-1}$$

$$= h'(c)\,[-1] \qquad = \dfrac{-\alpha}{(x+1)^{1+\alpha}}$$

$$= \alpha\,(c+1)^{-1-\alpha} \le h(k+1)^{-1-\alpha}$$

$$\le C'\, n^{-1-\alpha}.$$

some linear function of n. Hence, I can find the constant C prime such that this difference is upper bounded by C prime times n raise to minus 1 minus alpha. So, one can do some simple algebra here to explicitly get an expression for C prime and you know get some expression like this. So, what this implies is that if you look at the tail sum that is for k bigger than equal to floor of n over 2, this expression we have shown is upper bounded by some constant times n raised to minus 1 minus alpha and since again this expression does not depend on k, I can pull it outside this summation and we will now left with sum of ukn type terms and we have already shown that this ukn. This u k n we have already shown is upper bounded by e raised to minus 2 n minus 1 minus k alpha n minus 1. So, that expression I have written over here and hence we end up with some term like this.

So, in this expression, observe that we have e raised to minus 2 times alpha n minus 1, and the whole thing raised to n minus 1 minus k. So, if you think of this, this expression does not depend on k, and we are using this expression and summing over its powers where k is what is changing. So, what we have over here is a geometric series. Is this okay? And by making use of the properties of geometric series, what one can do is—I think there is some mistake over here—let me check this. So, this I will cancel off. Okay, so this expression that we have over here, actually, I have written it as it is, and this expression over here, right? I am making use of the fact that, you know, a If you had a geometric series, so let us say you had some expression like a m plus a m plus 1 plus dot dot dot, right?

So, this expression, whenever a is strictly less than 1, is actually upper bounded by a raised to m by 1 minus a, right? So, this is what we have. And in the same way, since we have a geometric series here, right, what I do is I take the smallest value of k, which is floor of n over 2, right, and that expression I write and divide it by 1 minus the, you know, the term with which the geometric series is defined. That is the, you know, A-type expression. The A-type expression is something like this, which is what I have written over here.

So, now what I will do is I will try to get a lower bound for this expression so that I can get an upper bound for this expression. So, towards that, we know that 1 minus x is less than e raised to minus x. So, this would imply that 1 minus e raised to minus x is less than or equal to x, right? So, this simple fact that we know will only give a lower bound—sorry, an upper bound—on 1 minus e raised to minus x. However, since this expression is in the denominator and we want an upper bound for the whole expression, it is preferable that we get a lower bound for the term that is there in the denominator. So, towards that What one can make use of is this inequality: that is, e raised to minus x is actually less than or equal to 1 minus x plus x squared over 2 for x greater than or equal to 0. So, one can actually show this thing: that for x greater than or equal to 0, e raised to minus x is less than 1 minus x plus x.

x squared over 2, right? And hence, if I put this e raised to minus x on the other side, one can see that 1 minus e raised to minus x is lower bounded by x minus x squared over 2, right? And for x less than or equal to 1, one can show that this expression is actually lower bounded by x over 2, which helps us conclude that A lower bound on 1 minus e raised to minus x is actually x over 2. And since we have something like this over here, we can show that this lower bound—I mean, a lower bound for this expression—is actually 2 alpha n minus 1 over 2, and the 2 cancels off, and we end up with alpha n minus 1. And hence, one can show that the sum of the last n over 2 terms is actually c prime times n raised to minus 1 minus alpha. Times e raised to the n over 2th term times alpha n minus 1 divided by alpha n minus 1.

$$u_k^{(n)} \leq e^{-2(n-1-k)\alpha_{n-1}}$$

this implies that

$$\sum_{k \geq \lfloor \frac{n}{2} \rfloor}^{n-1} (\alpha_k - \alpha_{k+1}) u_k^{(n)} \qquad \left(\frac{-2\alpha_{n-1}}{e}\right)^{n-1-k}$$

$$\leq c'n^{-1-\alpha} \sum_{k \geq \lfloor \frac{n}{2} \rfloor}^{n-1} e^{-2(n-1-k)\alpha_{n-1}}$$

$$\leq c'n^{1-\alpha} e^{\cdots}$$

$$= c' n^{-1-\alpha} \frac{e^{-2\left(n-1-\lfloor \frac{n}{2} \rfloor\right)\alpha_{n-1}}}{1 - e^{-2\alpha_{n-1}}}$$

For $x > 0$,

$$1 - x \leq e^{-x}$$
$$\Rightarrow 1 - e^{-x} \leq x$$

$$e^{-x} \leq 1 - x + \frac{x^2}{2}$$

Hence, $1 - e^{-x} \geq x - \frac{x^2}{2}$

Now, for $x \leq 1$,

$$x - \frac{x^2}{2} \geq \frac{x}{2}$$

Hence, $1 - e^{-x} \geq \frac{x}{2}$

$$\boxed{a^m + a^{m+1} + \cdots \leq \frac{a^m}{1-a}}$$

this implies that

$$\left(\sum_{k \geq \lfloor \frac{n}{2} \rfloor}^{n-1} (\alpha_k - \alpha_{k+1}) u_k^{(n)}\right)$$

$$\leq c'n^{-1-\alpha} \frac{e^{-2\left(n-1-\lfloor \frac{n}{2} \rfloor\right)\alpha_{n-1}}}{\alpha_{n-1}}$$

Since $n - 1 - \lfloor \frac{n}{2} \rfloor \geq \frac{n}{2} - 1$,

we have

$$\sum_{k \geq \lfloor \frac{n}{2} \rfloor}^{n-1} (\alpha_k - \alpha_{k+1}) u_k^{(n)}$$

$$\leq c'n^{-1} e^{-(n-2)\alpha_{n-1}}$$

$$= O\left(\frac{1}{n}\right)$$

So, this is the consequence of the algebra that we did here, right? And one can separately check that n minus 1 minus the floor of n over 2 is actually n over 2 minus 1. This is something very easy to check. And hence, one can show that this whole expression right here is upper bounded by C prime e raised to minus 1 minus alpha, right? Minus alpha, and in the denominator, you had alpha n minus 1, which is again n raised to minus alpha. So, this expression that we have

and this thing will cancel off, right? And this expression that we have separately—okay, this expression—we can show that it is upper bounded by e raised to minus n over 2 times alpha n minus 1, right? Now, this expression is upper bounded by 1, right? Because this term over here is non-negative, and this gives us a 1 over n expression. So again, one can conclude that the tail of this sum is also of order 1 over n. So, let me summarize what we have done so far. In the previous class, we showed that the expected value of theta n

squared is upper bounded by Sn times sigma squared, and this Sn we showed is upper bounded by an expression like this. where this g n minus 1 was of order alpha n minus 1, where the other expressions were little o of alpha n minus 1. And what we have done so far is to show that this sum—which, you know, was like together—so after we broke it up into the first n over 2 terms and the second n over 2 terms, we have shown that this expression is of order 1 over n, and the tail is also of order 1 over n. In other words, This sum and this sum, if you combine them, they are of order 1 over n, and I mean the reasons why they are of order 1 over n are different.

In the first case, we show that this term leads to, you know, an exponentially decaying bound, right? And that term is the reason why you have order 1 over n. Here, on the other hand, we showed that, you know, The total sum, okay, this sort of leads to a geometric sum, and this gk minus gk plus 1 is what leads to an order 1 over n type expression, right.

So, hence, one can conclude that the whole sum is of the form, okay. So, this whole thing is of the form sigma square, right. Right, alpha n minus 1 plus little o of alpha n minus 1 plus order of 1 over n, okay? So this is like big O, whereas this is little o, and we know that, you know, 1 over n decays faster than 1 over n raised to

So, this is something we know: that this decays faster than this. In other words, you know, for large enough n, 1 over n will be much less than 1 over n to the power alpha, and in particular, one can show that your 1 over n is actually little o of alpha n minus 1. Hence, whatever we had over here, we can combine all of them and say that this is little o of 1. Alpha n minus 1, and hence one can conclude that the final convergence rate is sigma square over 2 alpha n minus 1 plus some terms that decay faster. So, this brings us to the desired convergence rates, and let us, you know, quickly summarize what we have understood. So, we looked at the convergence rate of the mean estimation algorithm in two cases.

The first case is where alpha n equals 1 over n plus 1, and the second case is of the slowly decaying step size choice where alpha n equals 1 over n plus 1 to the power alpha. And what we have shown is that for alpha strictly less than 1, the convergence rate is actually poorer compared to the rate when alpha n equals 1 over n plus 1. So, this suggests if your goal is to work with—I mean, if your main criteria is convergence rate—then the 1 over n plus 1 step size will be, in some sense, better. Now, one can ask, okay. If 1 over n plus 1 is the best step size choice, then if you go back to that mean estimation algorithm, we

observed that your xn plus 1, that is the algorithm's estimated time n plus 1, little xn plus 1, right?

That was just a simple sample mean, right? So, then one can ask, okay, if such a simple sample mean maybe gives, you know, a better convergence rate. Maybe there could be some more sophisticated function of these random variables that could lead to a better convergence rate, right? So, in the next class, we will show that that also will not lead to a better convergence rate. In fact, the best convergence rate in terms of order that one can obtain is indeed 1 over n, and we will prove this in the next class, which will help us conclude that, at least in the context of stochastic approximation algorithms, from the point of view of convergence rates, choosing this alpha n equals 1 over n plus 1 is ideal, right. So, with this, let me stop today's lecture. Hope I can meet you in the next class.

Thank you.